AAV-deliverable hypercompact adenine base editors based on transposase B guided by engineered RNA

Do Yon Kim
GenKOre
Yuhee Chung
GenKOre
Yujin Lee
KRIBB
Dongmin Jeong
KRIBB
Kwang-Hyun Park
KRIBB
Hyun Jung Chin
KRIBB
Jeong Mi Lee
KRIBB
Seyeon Park
GenKOre
Sumin Ko
GenKOre
Jeong-Heon Ko
KRIBB
Yong-Sam Kim (yskim@genkore.com)
GenKOre

Article

Keywords: AAV, Adenine base editing, TnpB, Gene therapy, Precise genome editing

Posted Date: April 12th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1326630/v1

Version of Record: A version of this preprint was published at Nature Chemical Biology on August 1st, 2022. See the published version at https://doi.org/10.1038/s41589-022-01077-5.

Abstract

Transposon-associated transposase B (TnpB) is deemed an ancestral protein for type V, Cas12 family members, and the closest ancestor to UnCas12f1 due to its high sequence similarity. Previously, we reported a set of engineered guide RNAs supporting high indel efficiency for Cas12f1 in human cells. Here, we suggest a new technology whereby the engineered gRNAs also manifest highly efficiency programmable endonuclease activity for TnpB, termed TaRGET (TnpB-augment RNA-based Genome Editing Technology). Having this feature in mind, we established TnpB-based adenine base editors. A codon-optimized Tad-Tad mutant (V106W, D108Q) dimer fused to the C-terminus of dTnpB (D354A) showed the highest levels of A-to-G conversion. The limited targetable sites for TaRGET-ABE were expanded by either developing several PAM variants of TnpB or by engineering TnpB and optimizing deaminases at PAM-distal and PAM-proximal sites, respectively. When delivered by AAV, the TaRGET-ABE showed potent A-to-G conversion rates in human cells. Collectively, the TaRGET-ABE will contribute to improving precise genome-editing tools that can be delivered by AAV, thereby harnessing the development of CRISPR-based gene therapy.

Main

Living organisms are subjected to spontaneous genetic variations, which form the basis for biodiversity and evolution. The random nature of genetic variation as well as non-biological factors are also responsible for a variety of genetic disorders. Of the types of genetic variations identified in humans, single-nucleotide variations (SNVs) account for nearly half of disease-related mutations[1]. This suggests that the development of site-specific, precise genome editing tools holds promise for the treatment of otherwise intractable genetic disorders. CRISPR/Cas-mediated base-editing systems have addressed the unsatisfactory modification efficiency of homology-directed repair (HDR) by exploiting the high gene-targeting capability of the CRISPR/Cas systems. The catalytically inactive Cas (dCas) or nickase Cas (nCas) fused to naturally occurring or engineered deaminases led to highly efficient single-nucleotide alterations including the C:G-to-T:A[2] and A:T-to-G:C[3] conversion. The introduction of an *E. coli*-derived uracil DNA N-glycosylase also enabled C:G-to-G:C transversion[4]. Moreover, the base editing systems guarantee higher levels of safety from a clinical view point because they enable precise genome editing with negligible or low levels of double-strand breaks (DSBs). In addition to such innate high efficiency and safety, base editing systems have further evolved in various aspects including decreased off-target editing, enhanced conversion specificity, broadened editing windows, increased editing efficiency, and control of unwanted editing[5-7].

Despite the dramatic improvements of base-editing system *per se*, the challenges on the delivery side remain a major hurdle preventing base editors from being widely used in clinical applications[8]. Adeno-associated viruses (AAVs) are considered as a validated delivery platform due to their relatively high delivery efficiency, low immunogenicity, relatively low level of concern over chromosomal integration, and their other useful and clinically validated features. Moreover, the identification of and engineering efforts

for various AAV serotypes has expanded the range of targetable tissues or organs[9, 10]. Nonetheless, the multifaceted merits of AAVs are drastically compromised by their low packaging capacity of 4.7 kb. The limited cargo sizes are even more problematic for base-editing systems due to heavy Cas proteins and an additional deaminase protein[8, 11]. Even the adoption of a light Cas effector such as SaCas9 (3.16 kb) does not alleviate the payload limit issue of AAVs when additional essential components, in this case a promoter, one or more guide RNA cassettes and a poly-adenylation signal sequence are deployed in a vector. Although the oversize issue was, in part, addressed by the split two-vector system[12], this led to other related challenges for clinical applications; the manufacture of AAV particles on a good manufacturing practice (GMP) basis is costly, and the need to operate two different manufacturing lines can be burdensome. Furthermore, recent evidence indicates that higher dosages of AAV may induce liver toxicity[13]. Therefore, it is crucial to develop an AAV-packable, but efficient base editor from a clinical perspective.

Recently, hypercompact programmable RNA-guided DNA nucleases have been suggested as a promising option for more versatile clinical applications[14-16]. In line with this research trend, we transformed the non-functional CRISPR-Cas12f1 system into a highly efficient genome editor through the extensive engineering of gRNA[14]. Here, we show that our engineered RNAs can be used as similarly efficient guide RNAs for transposon-associated transposase B (TnpB), an evolutionarily direct ancestor for UnCas12f1, suggesting a TaRGET (TnpB/augment RNA-based Genome Editing Technology) system. More importantly, catalytically inactive TnpB (dTnpB) fused to optimally configured deaminases offers a robust platform for an AAV-deliverable adenine base editing system. We also expanded the targetable range by developing PAM variants and a protospacer-protruding mutant of TnpB and using a unique combination of deaminases. This TaRGET-ABE system is expected to contribute to providing an improved tool set for the clinical applications.

Results

TaRGET (Transposase B/augment RNA-based Genome Editing Technology) system

Recently, several hypercompact ancestral genes have been reported to show programmable RNA-guided endonuclease activity. In particular, the IS200/IS605 transposon-encoded TnpB has RuvC-like domain can be used as a genome editor when complexed with a compatible gRNA[17]. Karvelis *et al*. showed that TnpB of *Deinococcus radiodurans* ISDra2 (ISDra2TnpB) is a programmable endonuclease that is guided by right element (RE)-derived RNA (reRNA) to cleave DNA next to the 5'-TTGAT transposon associated motif (TAM)[18]. TnpB from *A. macrosporangiidus* also showed an omega gRNA-specific dsDNA cleavage *in vitro* with the TCAC TAM preference, though the *in vivo* indel efficiency was not explored[18].

Type V Cas proteins, namely Cas12 family members, are likely to evolve from TnpB, and UnCas12f1 is an early member of TnpB-origin Cas effectors[19]. Because the IS200/IS605 family element transposase accessory protein TnpB from the *Candidatus Woesearchaeota* archaeon (hereafter TnpB) shares a

perfectly matched nucleotide sequence with UnCas12f1 except for 5'-terminal 28 amino acid residues (Supplementary Table 1), the TnpB endonuclease may share certain molecular properties with it, including endonuclease activity and RNA binding. To test this possibility, the RNA-guided programmable nuclease activity was investigated for TnpB. In a previous study[14], we developed several engineered versions of sgRNA for Cas12f, including ge3.0, 4.0, and 4.1. Identical to Cas12f, TnpB showed no indel-formation activity with canonical gRNA in HEK293T cells. However, TnpB exhibited significantly increased indel activity with the engineered (augment) gRNAs (Supplementary Fig. 1). Despite the slight target-dependent difference in the indel levels between Cas12f and TnpB, the overall cleavage showed a nearly identical pattern, indicating the orthogonal use of the augment RNAs for TnpB (TaRGET). Because the base editing efficiency usually depends on the indel efficiency of wild-type Cas effector proteins, it is necessary to start with a Cas system that shows sufficiently high indel efficiency. Thus, we compared the indel efficiencies of TaRGET, ISDra2TnpB, and AmaTnpB at the PCSK9 *loci* in HEK293T cells. The target sites are not exactly shared due to differences in respective PAM sequences. Therefore, 11 sites between exon 5 and exon 8 were selected instead in the PCSK9 genome sequence. Interestingly, the TaRGET system showed significantly higher indel efficiency, compared to those of ISDra2TnpB and AmaTnpB (Fig. 1a). Therefore, we concluded that the TaRGET system is feasible for use as a platform for the development of a compact base editing system.

Feasibility of dTnpB-based adenine base editors

Based on the information on the residues involved in the catalytic activity for Cas12f[20, 21], we constructed four catalytically inactive mutants of TnpB (D354A, E450A, R518A, and D538A). Each mutant was tested as to whether the DNA cleavage activity was completely eradicated while allowing the preservation of the gene-targeting capability. An *In vitro* DNA digestion assay and an indel assay in HEK293T cells revealed that all of the mutants had null endonuclease activity (Supplementary Fig. 2), and we selected a dTnpB (D538A) mutant based on a previous CRISPRa experiment[14].

A size-exclusion chromatogram was used to estimate the molecular mass of sgRNA-bound TnpB as *ca*. 194 kDa, suggesting that TnpB formed a homo-dimer in the presence of engineered gRNA, similar to UnCas12f1[20, 21] (Fig. 1b). In this case, the orientation of deaminase fusion may affect the base-editing property. To test this possibility, we constructed TnpB-based adenine base editors by fusing the wild-type Tad-mutant Tad (Tad-Tad*) or Tad*-Tad to either the N- or C-terminus of dTnpB[3] (Fig. 1c). These constructs were tested for two validated targets[14], one showing an A-rich sequence at the PAM-proximal region and the other showing this sequence at the PAM-distal region (Fig. 1d). The deaminase architectures fused to the C-terminal orientation (ABE-C1 and ABE-C2) showed substantial levels of A-to-G conversion activity, whereas the N-terminally fused modules (ABE-N1 and ABE-N2) were only marginal in terms of conversion activity. Conversions were only observed at the PAM-proximal regions. Because ABE-C2 showed higher conversion rates compared to the other ABEs, it was used to identify a base-editing window. In an experiment where two PAM-proximal A-rich sequences were targeted, a conversion was only observed within the window of A2 to A5, with the most prominent conversion activity observed at

positions A3 and A4 (Fig. 1e). Base editors evolved with various engineered versions of Tad[22]. Thus, we compared the A3-to-G3 conversion rates for all Tad variants developed thus far and found that the architectures of the codon-optimized Tad-Tad* (V106W, D108Q)[23,24] showed the highest conversion rates compared to any of the other forms (Supplementary Fig. 3a, Supplementary Table 1). We presented the optimized Tad dimer as Tad-Tad** and designated this ABE form as TaRGET-ABE-C3. The length of the linker and the position of NLS did not affect the base editing efficiency (Supplementary Fig. 3b). TaRGET-ABE-C3 was compared with several adenine base editors with respect to the editing windows and conversion efficiency (Fig. 1f). As mentioned above, the base editing window of TaRGET-ABE-C3 was formed at a relatively narrow range, similar to the recently reported ABEMINI[15]. However, the overall conversion efficiency of TaRGET-ABE-C3 was significantly higher than that of Cas12f-based ABEMINI, though it was lower than those of SpCas9 nickase-based ABEs, such as ABE7.10, ABE8e, and ABE9 (Fig. 1g).

In a previous study, we presented three different versions of sgRNA for UnCas12f1, ge3.0, ge4.0, and ge4.1, indicating that the selected sgRNA version would affect the base-editing efficiency. To investigate this possibility, we selected 18 targets that show different indel activity outcomes depending on the sgRNA version. Similar to earlier work, 15 out of 18 sites showed the correlation between the gRNA version and indel/conversion efficiency (Fig. 1h). That is, the selection of the most suitable gRNA must come first for the most desirable base-editing outcome. Taken together, TnpB-ABE-C3 guided by an optimal gRNA version shows optimal base-editing performance.

Expanding targetable sites via TnpB and Tad engineering

Identical to Cas12f1, the *in vitro* cleavage assay indicated that TnpB showed a PAM preference for TTTR (TTTA and TTTG), which means that targetable sites are quite restricted (Fig. 2a). Thus, we attempted to develop TnpB mutants with preference to non-TTTR PAM and to apply PAM variants to a wider range of sites. To do this, we initially constructed a PAM library vector, which was achieved by securing individual PAM clones (4^4=256 clones) and then mixing them at an equal molar ratio to ensure even distributions of each. The PAM library vectors were digested with sgRNA ge_4.1 and the TnpB PAM variant proteins. The cleaved vectors were amplified by adaptor ligation and PCR. A deep sequencing analysis enabled PAM variant-PAM preference matching (Fig. 2b). To select PAM variants with retained dsDNA cleavage activity levels, we prepared different HEK293T clones each carrying different PAM mutants at an NLRC4 *locus* via homology-directed repair. (For details, please refer to Supplementary Fig. 4) This approach would make it possible to compare the relative indel efficiency of each PAM variant, compared to the wild-type TnpB.

Because TnpB shows an identical PAM preference and shows sequence conservation in the DNA-binding region with UnCas12f1, we selected candidate amino acids on TnpB based on the structural characterization of UnCas12f1[20, 21], in this case, S170, Y174, A184, S188, R191, Q225, Y230, V271, and Q272. Each candidate site was mutated into all possible 19 amino acids, and each PAM variant candidate was tested with respect to the dsDNA cleavage activity *in vitro* for the altered PAMs as described in the scheme of Fig. 2b. The PAM-variant candidates were selected for the criteria of 1) high

total sequencing reads and 2) a high sequencing read ratio for a specific PAM. The *in vitro* cleavage and the deep sequencing analysis enabled the screening of PAM-variant candidates (Supplementary Fig. 5). The results indicate that several variants showed redundantly identical PAM preference. For instance, the S170T, S188Q, S188H, Q225T, Q225F, and Q272K variants showed a high TGTA PAM preference. Among the variants, the S188Q variant showed highest indel frequency for TGTA PAM compared to the other variants, when tested in PAM sequence-altered HEK293T cells as described in Supplementary Fig. 4. Likewise, the S188Q, S188K, and R191K mutants showed high indel frequencies for TCTG, TGTG, and TTTC PAMs, respectively (Fig. 2c). The S118K variant showed broader PAM specificity, in this case TTTT and TTTC as well as TTTA and TTTG, i.e., TTTN (Fig. 2d). To test the application of the PAM variants to adenine base editing in a non-TTTR PAM context, the A-to-G conversion activity of the PAM variants were tested for different sites with altered PAMs. As shown in Fig. 2e, each PAM variant showed different levels of A3-to-G3, A4-to-G4, and A5-to-G5 conversion activities. A suitable PAM variant can be selectively used for a specific sequence context, or a variant showing a multi-PAM preference such as the S188K variant can be deployed for multiplexed targeting. Collectively, the engineering of TnpB expanded the occupancy of targetable base-editing sites from 0.78% to 3.12%.

Despite the expansion of targetable sites using PAM variants, the editable incidence is still limited because a prominent editing window formed at positions 3 and 4 (This feature is occasionally favorable for specific editing) Expanding or shifting the window could be an additional option by which to expand the applicability of the TaRGET-ABE system. Structural modeling of the TnpB-gRNA ribonucleoprotein complex identified possible mutation sites at Ile^{159} and Ser^{164}. The model indicates that the bases at position of 5 and 6 are concealed in a pocket of the WED domain (Supplementary Fig. 6). We speculated that the replacement of Ile^{159} and Ser^{164} with a bulky amino acid would make the bases of positions 5 and 6 more protruding (Fig. 2f), which would make deaminases more accessible to those bases. We created the I159W and S164Y mutants and applied them to adenine base editing for several targets, each of which carrying A at a different position. When we compared the editing efficiency of the variants with that of wild-type TaRGET-ABE-C3, the S164Y mutant led to a dramatic compromise in the A-to-G conversion rates at positions 3 and 4 without window expansion. However, the I159W mutant upheld the conversion rates at the positions 5 and 6 with retained A3 and A4 conversions (Fig. 2g). The last approach was related to a divergent architecture of the deaminase module. While constructing various combinations of Tad variants, we fortuitously found that dTnpB-Tad-eTad modules showed a window expansion at position 2. The eTad sequence was originally used as a monomeric deaminase for the ABE8e version[25]. The fusion of the Tad-Tad8e dimer module to dTnpB (D538A), hereafter referred to as TaRGET-ABE-C3.1, induced dramatically boosted conversion at position 2 with sustained conversion efficiency outcomes at positions 3 and 4 (Fig. 2h).

We attempted to validate the TaRGET-ABE-C3.1 system for 25 endogenous sites (Fig. 2i, Supplementary Fig. 7 and Supplementary Table 2). The distribution confirmed the most prominent base editing at positions 2-5 without the application of the I159W mutation. Nonetheless, we identified two sites for which positions 17 and/or 18 was edited at a relatively high efficiency rate. Thus, this non-canonical

editing should be monitored on a per-site basis. Taken together, the engineering and reconstruction of TnpB and Tad modules largely broadened the otherwise highly restricted base editing range by both expanding targetable PAMs and shifting or expanding the base editing windows.

Adenine base editing via AAV delivery

nSpCas9 (D10A)-based adenine base editors enable highly efficient A-to-G conversions in eukaryotic cells when delivered by plasmid vectors[3, 25]. However, the AAV delivery is limited due to the oversized deaminase-dCas9 modules[8]. This limitation can be overcome by using a split-AAV vector delivery[12] or miniABE8e[26]. While all of the engineered ABEs compromised the full activity of Cas9-based ABE systems, our TaRGET-ABE system is sufficiently compact such that it can be delivered in an all-in-one AAV vector. Furthermore, there remains space for additional cargo within a payload size limit of ~4.7 kb. One of the applications utilizing the additional cargo space would be multiplexed base editing. We produced AAV2 particles where the TaRGET-ABE-C3 system was charged with one sgRNA (sgRNA 1 for site3 or sgRNA2 for site5) or paired sgRNAs simultaneously targeting site3 and site5 (Fig. 3a). HEK293 cells were transduced at a multiplicity of infection (MOI) of 100,000 for vector systems for ten days, during which the cells were sub-cultured upon five day post initial transduction and the MOI was kept constant through additional treatments of AAV particles with fresh media. When a single sgRNA was loaded onto the AAV vector, a target-specific base editing was achieved (Fig. 3b). Interestingly, we were able to perform multiplexed A-to-G conversions using paired sgRNAs in a single AAV particle. Moreover, the conversion efficiency at each site obtained using the paired gRNA-AAV particles was not compromised, compared to those obtained using one sgRNA-charged AAV particle.

The capability of charging paired sgRNAs in an all-in-one AAV vector system can act as a ‘double-edged sword’ regarding substitution-based treatments of certain diseases. We illustrate this concept for a possible treatment strategy for cancer. Epidermal growth factor receptor 4 (ErbB4; HER4) is a kinase that stimulates oncogenesis and cancer progression in many cancer types, and chemical or biological inhibitors are used for the treatment of cancer[27]. We loaded two sets of sgRNAs together with TarGET-ABE-C3 in an all-in-one AAV vector. sgRNA1 aims to induce exon skipping and a frame-shift mutation by substituting a splicing acceptor consensus sequence (-AG-) with a splicing-skipping sequence (-GG-). sgRNA2 induces the skipping of the exon involved in the binding of growth factors. The use of either one of the two sgRNAs can produce non-functional receptors, but the concomitant charge of the two sgRNAs can further increase the frequencies of the occurrence of non-functional ErbB receptors (Fig. 3c). We screened two intron-exon interface targets that meet the requirements of PAM and the reading frame: One (sgRNA1) is at the intron I-exon 2 and the other (sgRNA2) at intron III-exon 4. AAV2 particles carrying either one of the two sgRNAs or both were produced in HEK293T cells and treated in H661 cells at an MOI of 10^5. TaRGET-ABE-C3 carrying sgRNA1 or sgRNA2 produced non-functional mRNAs with at the frequency of 17.3±2.5% and 13.2±2.3%, respectively. However, the percentage of non-functional mRNAs increased to 26.8±3.7% for AAV particles carrying both sgRNAs (Fig. 3d). This ‘double-edged sword’ effect was manifested in the stalling of the growth of cancer cells, where the two sgRNAs collaborated to retard

the growth of ErbB4-positive A549 cells (Fig. 3e). Taken together, these results suggest that the hypercompact TnpB-based adenine base editors provide a useful and precise genome editing tool delivered by AAV. It is important to note that a more universal PAM variant would expand the application of the exon-skipping strategy for gene knockout, particularly genes consisting of a few exons, including the transthyretin (TTR)[28] and proprotein convertase subtilisin/kexin type 9 (PCSK9)[29,30].

Discussion

Precise genome editing would provide a radical treatment option for various genetic disorders, most of which rely on symptomatic therapies without a curative regimen. The base-editing strategy is assumed to bring significant clinical benefits for potential patients with a single substitution mutation. Recently, an adenine-base editor was proposed for safer gene knockout without double-stranded DNA breaks through an exon-skipping strategy[29, 30]. Despite these wider applications to genetic manipulations *in vivo*, the clinical utility of adenine-base editors has been largely restricted due to the delivery failure when using AAV particles. Nguyen *et al.*, made an arduous effort to deliver an nSaCas9-based adenine base editor in the all-in-one AAV vector[26], but the conversion efficiency was quite low (less than 1%). Lipid nanoparticles are considered as a delivery system for these heavyweight CRISPR editors, but clinically approved ones are confined to the targeting of the liver tissue[28]. Thus, the development of hypercompact TaRGET-ABE would expand the spectrum of precise genome editing-based treatments. The TaRGET-ABE system would also make the development of gene therapy applicable to a wider range of genetic diseases.

The property of persistency related to AAV-assisted gene delivery provokes safety concerns over *in vivo* gene therapy. Because the TaRGET-ABE system adopts catalytically dead TnpB instead of a nickase, the issue of a residual level of indel mutations that are occasionally observed for nickase Cas proteins can be excluded. However, the issue of guide RNA-independent as well as –dependent off-target activity remains[5]. Our previous study indicated less tolerance to mismatches and a low incidence of off-targets for the CRISPR-Cas12f1 system, but the specificity of the TaRGET system remains to be elucidated. Furthermore, gRNA-independent off-target concerns must be addressed and, further, TaRGET-specific strategies to mitigate those concerns need to be developed.

We attempted to develop a cytosine base editor using dTnpB by employing various cytidine deaminases validated beforehand. However, we found that the C-to-T conversion efficiency was not satisfactory (in fact, <5% for, at least, tested targets) in HEK293T cells. This low efficiency may arise from the accessibility of fused deaminases to nucleotides within editing windows possibility due to incompatible orientation or any structural hindrance. A scrutinized design of deaminase modules based on structural modeling would enable one to render a functional TaRGET-CBE system. Furthermore, the development of a nickase TnpB, which could be obtained possibly by an alternative strategy to the previous sequence manipulations inside the RuvC domain, might allow one to establish a prime editing system that can be delivered in an all-in-one AAV vector system.

Declarations

Acknowledgments

This work was supported by a grant through the 'KRIBB Research Initiative Program (KGM5382113)', and the National Research Foundation of Korea (INNOPOLIS) grant funded by Ministry of Science and ICT (2021-DD-RD-0178-01), and by an 'Alchemist Project' funded by the Ministry of Trade, Industry and Energy (20012445).

Author contributions

Y.-S.K. conceived the study and designed the experiments. D.Y.K. and Y.C. performed overall experiments. D.J and Y.L. constructed the PAM library vectors and PAM variants of TnpB. H.J.C. and J.M.L performed base-editing assays in cells. S.P. and S.K. derived PAM-mutant HEK293T cells. K.-H.P performed the structural analysis for TnpB engineering. J.-H.K. and Y.-S.K. interpreted data. Y.-S.K. wrote the manuscript.

Competing interests

The authors declare competing financial interests. The authors through GenKOre have filed patent applications on the TaRGET-ABE and PAM variants of TnpB. Y.-S.K. and D.Y.K. are co-founders of GenKOre. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The authors declare no competing non-financial interests.

References

1. Anzalone, A.V. et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149-157 (2019).
2. Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A., Liu, D.R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533, 420-424 (2016).
3. Gaudelli, N.M. et al. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* 551, 464-471 (2017).
4. Kurt, I.C. et al. CRISPR C-to-G base editors for inducing targeted DNA transversions in human cells. *Nat Biotechnol* 39, 41-46 (2021).
5. Porto, E. M. et al. Base editing: advances and therapeutic opportunities. *Nat Rev Drug Discov* 19, 839-859 (2020).
6. Naeem, M., Majeed, S., Hoque, M. Z., Ahmad, I. Latest Developed Strategies to Minimize the Off-Target Effects in CRISPR-Cas-Mediated Genome Editing. *Cells* 9(7), 1608 (2020).
7. Molla, K. A., Yang, Y. CRISPR/Cas-Mediated Base Editing: Technical Considerations and Practical Applications. *Trends Biotechnol* 37, 1121-1142 (2019).

8. Stevanovic, M., Piotter, E., McClements, M. E., MacLaren, R. E. CRISPR Systems Suitable for Single AAV Vector Delivery. *Curr Gene Ther* 21, https://doi.org/10.2174/1566523221666211006120355 (2021).
9. Wang, D., Tai, P.W.L. & Gao, G. Adeno-associated virus vector as a platform for gene therapy delivery. *Nat Rev Drug Discov* 18, 358–378 (2019).
10. Li, C., Samulski, R.J. Engineering adeno-associated virus vectors for gene therapy. *Nat Rev Genet* 21, 255–272 (2020).
11. Wu, Z., Yang, H., Colosi, P. Effect of genome size on AAV vector packaging. Mol Ther 18 (1), 80-86 (2010).
12. Levy, J.M. et al. Cytosine and adenine base editing of the brain, liver, retina, heart and skeletal muscle of mice via adeno-associated viruses. *Nat Biomed Eng* 4, 97–110 (2020).
13. Chand, D. et al. Hepatotoxicity following administration of onasemnogene abeparvovec (AVXS-101) for the treatment of spinal muscular atrophy. J Hepatol 74(3), 560-566 (2021)
14. Kim, D.Y. et al. Efficient CRISPR editing with a hypercompact Cas12f1 and engineered guide RNAs delivered by adeno-associated virus. *Nat Biotechnol* (2021).
15. Xu, X. et al. Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing. *Mol Cell* 81(20), 4333-4345 (2021).
16. Wu, Z. et al. Programmed genome editing by a miniature CRISPR-Cas12f nuclease. *Nat Chem Biol* 17, 1132–1138 (2021).
17. Altae-Tran, H.et al. The widespread IS200/IS605 transposon family encodes diverse programmable RNA-guided endonucleases. Science 374, 57-65 (2021).
18. Karvelis, T. et al. Transposon-associated TnpB is a programmable RNA-guided DNA endonuclease. *Nature* 599, 692–696 (2021).
19. Shmakov, S. et al. Diversity and evolution of class 2 CRISPR–Cas systems. *Nat Rev Microbiol* 15, 169–182 (2017). .
20. Takeda, S. N. et al. Structure of the miniature type V-F CRISPR-Cas effector enzyme. *Mol Cell* 81(3), 558-570 (2021).
21. Xiao, R., Li, Z., Wang, S., Han, R., Chang, L. Structural basis for substrate recognition and cleavage by the dimerization-dependent CRISPR-Cas12f nuclease. *Nucleic Acids Res* 49, 4120-4128 (2021).
22. Molla, K.A. et al. Precise plant genome editing using base editors and prime editors. *Nat. Plants* 7, 1166–1187 (2021).
23. Rees, H.A., Wilson, C., Dorman, J.L., Liu, D.R. Analysis and minimization of cellular RNA editing by DNA adenine base editors. *Sci. Adv.* **8**, eaax5717.
24. Jeong, Y.K. et al., Adenine base editor engineering reduces editing of bystander cytosines. *Nat. Biotechnol.* **39**, 1426-1433 (2021).
25. Richter, M.F. et al. Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nat Biotechnol* 38, 883–891 (2020).

26. Nguyen Tran, M.T. et al. Engineering domain-inlaid SaCas9 adenine base editors with reduced RNA off-targets and increased on-target DNA editing. *Nat Commun* 11, 4871 (2020).
27. Sharma, B., Singh, V. J., Chawla, P. A. Epidermal growth factor receptor inhibitors as potential anticancer agents: An update of recent progress. *Bioorg Chem* 116, 105393 (2021).
28. Gillmore, J. D. et al. CRISPR-Cas9 In Vivo Gene Editing for Transthyretin Amyloidosis. *N Engl J Med* 385, 463-502 (2021).
29. Rothgangl, T. et al. In vivo adenine base editing of *PCSK9* in macaques reduces LDL cholesterol levels. *Nat Biotechnol* 39, 949–957 (2021).
30. Musunuru, K. et al. In vivo CRISPR base editing of *PCSK9* durably lowers cholesterol in primates. *Nature* 593, 429–434 (2021).

Methods

Plasmid vector construction

Human codon-optimized TnpB gene (Supplementary Table 1) was synthesized and cloned into a pCas12f-2A-EGFP vector (Addgene) by replacing the TnpB-coding sequence with Cas12f1 sequence. Various versions of Tad sequences were then fused at either 5’- or 3’-region of TnpB with 10-40 aa long linkers using NEBuilder HiFi DNA assembly master mix (New England BioLabs). Guide RNA sequences were positioned under U6 promoter 5’-upstream of CMV promoter using MluI restriction enzyme. Spacer sequences were cloned into by digesting the vectors with BbsI restriction enzyme for 1 h at 37°C. The catalytically dead TnpB were generated by site-directed mutagenesis by primers incorporating the intended base changes. Plasmid vectors for cell transfection were prepared using Nucleobond Xtra midi (MACHEREY-NEGEL). All vector constructs were sequence-verified using Sanger sequencing.

TnpB engineering

The construction of TnpB PAM variants and mutants (I159W and S164Y) was performed by site-directed mutagenesis. PCR amplifications were performed using Q5 Hot Start high-fidelity DNA polymerase (NEB) and the PCR products were ligated using KLD Enzyme Mix (NEB). The ligated products were transformed into DH5α *E. coli* cells. Mutagenesis was confirmed by Sanger sequencing analysis. The modified plasmid vectors were purified using a NucleoBond ® Xtra Midi EF kit (MN).

PAM library construction and PAM preference determination

Oligonucleotides harboring a protospacer (5’-CACACACACAGTGGGCTACC-3’) and PAM library sequence (NNNN) were synthesized (Bionics) and cloned into a PUC19 vector using an All in One^{TM} PCR cloning kit (Biofact). Each cloned vector was used to transform DH5α *E. coli* cells using an electroporator (Bio-Rad). Each transformant colony was grown at 37°C in LB broth until the culture reached an optical density of 0.6. Cells were collected by centrifugation at 3,500 g for 15 min. Plasmid vectors were prepared using a plasmid preparation kit (Biofact). Sequence was verified using Sanger sequencing analysis. Each vector

was spectrophotometrically quantified at 265 nm and mixed at an equal molar ratio to prepare 256 PAM library vectors. Plasmid vectors encoding TnpB PAM variants were used to transform BL21(DE3) *E. coli* cells. Each transformant colony was grown at 37°C in LB broth until the culture reached an optical density of 0.6. Cells were incubated at 18°C overnight in the presence of 0.1 mM isopropylthio-β-D-galactoside and then collected by centrifugation at 3,500g for 15 min. Cells were resuspended in 20 mM Tris-HCl (pH 7.6), 500 mM NaCl, 5 mM β-mercaptoethanol, 5% glycerol. Cell lysates were prepared by sonication followed by centrifugation at 15,000g for 15min. Guide RNA was synthesized using T7 RNA polymerase (NEB) in the presence of 1 μg of the purified plasmid and 4 mM NTPs (Jena Bioscience), purified using a Monarch® RNA cleanup kit (NEB) and aliquoted into cryogenic vials prior to storage in liquid nitrogen. Cell lysates (10 μl), guide RNA (1 μg), and PAM library plasmid vector (1 μg) were mixed at a final mixture 100 μl in 5mM Tris-HCl(pH7.5),25mM NaCl, 5mM $MgCl_2$, 1mM DTT buffer and incubated at 37°C for 2 h. The incubated samples were end-repaired using an NEBNext® Ultra™ II End Repair/dA-Tailing Module(NEB#E7546) at 20°C for 30min, and the reaction was terminated by incubating at 65°C for 30 min. After treated with RNase A at 100 mg/ml at room temperature for 15 min, plasmid DNA was purified using a HiGene^{TM} Gel&PCR purification kit (Biofact). The purified DNAs (200 ng) were ligated with an adaptor DNA (5'-AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC-3')(200 ng) using a LigaFast^{TM} Rapid DNA ligation kit (Promega). DNAs were PCR-amplified using a forward primer (5'-GTAAAACGACGGCCAGT-3') and a reverse primer (5'-GTGACTGGAGTTC-3') using a KOD One^{TM} PCR Master Mix (TOYOBO). The resulting PCR amplicons were labeled with Illumina TruSeq HT dual indexes. The final PCR products were subjected to 150-bp paired-end sequencing using an Illumina iSeq 100.

PAM-mutant cell lines

HEK293T cells (LentX-293T, Takara) were maintained in Dulbecco's Modified Eagle Medium (Corning) supplemented with 10% heat-inactivated fetal bovine serum (VWR) and 1% Penicillin-Streptomycin (WELGENE) in an incubator (37°C, 5% CO_2 atmosphere). PAM-varied oligonucleotides (90-mer) were synthesized as donor DNA. For transfection, 4 μg of SpCas9 plasmid vector was transfected with 4 μg of donor DNAs into $4X10^5$ HEK293T cells using a Neon transfection system (Invitrogen). The electroporation conditions were as follows: 1,300 V, 20 mA, 2 pulses. Following 3 days after transfection, single cell was placed in each well of 24-well plates (Corning) and grown for 3 weeks. Genomic DNA was prepared from each colony cells using a PureHelix^{TM} genomic DNA preparation kit (NanoHelix). PAM-containing region was amplified using KOD One^{TM} PCR Master Mix (TOYOBO) according to the manufacturer's instructions. The PAM sequence was verified by deep sequencing using an Illumina iSeq 100.

Measurement of substitution efficiency

HEK293T cells were transfected with vectors using a lipofection method. HEK293T or H661 cells were seeded into 24-well plate at a density of 1.0×10^5 /well 1 day before transfection. Six microliters of FuGene reagents (Promega) were mixed with 1.5 μg TnpB-ABE vector plus 500 ng gRNA-encoding PCR amplicon in 300 μl of Opti-MEM and incubated at room temperature for 15 min. The mixtures were added

to each well, and cells were grown for 3 to 5 days at 37°C and 5% CO_2. Genomic DNA was extracted by cell lysis with Martin's solution (50 mM Tris-HCl pH 8.5, 1 mM EDTA, 0.005% SDS, proteinase K). Samples for deep sequencing analysis were prepared by three rounds of PCR amplifications. For the primary PCR, 1 µl of cell lysis was amplified by target-specific primers designed to amplify targeted locus in 10 µl total volume. One µl of primary PCR products were amplified by primers with Illumina adapter sequence to produce 150 bp-long amplicons. Finally, Illumina TruSeq HT dual indexes were labeled on the PCR amplicons by PCR reactions. All PCR reactions were performed using KOD one PCR master mix (Toyobo) according to the manufacturer's instructions. Pooled amplicons were column-purified using a PCR purification kit (BioFact). The final PCR products were subjected to 150-bp paired-end sequencing using an Illumina iSeq 100. Indel frequencies were calculated by MAUND, which is available at https://github.com/ibs-cge/maund. The tested sites were compiled in Supplementary Table 3.

Production of adeno-associated virus

AAVpro 293T cells (Takara) were seed at the density of 2×10^7 onto a triple flask with a dimension of 500 cm^2 (Thermo) prior to transfection. Vectors encoding base editor components, pHelper, and pAAV-RC2/2 (or pAAV-RC2/9) were mixed at 1:1:1 molar ratio, and the vector mixture (150 µg) was used to transfect AAVpro 293T cells at the confluency of ~70% using a Polyethylenimine (Polysciences) transfection reagent according to the manufacturer's instruction. After 3 days post transfection, cells were harvested and lysed by three cycles of freezing and thawing. Cell lysates were treated with DNase I (Enzynomics) at the final concentration of 10 units/µl at 37°C for 30 min. A gradient solution was prepared by adding 4 ml of 60% iodixanol solution, 4 ml of 40% iodixanol, 5 ml of 25% iodixanol, and 7 ml of 15% iodixanol sequentially into a tube. Lysed samples were poured with care on top of the gradient solution. The gradient mixture was centrifuged in a 50.2 Ti ultracentrifuge rotor (Beckman) at 50,000 rpm at 10°C for 2.5 h. The viral fraction was retrieved using an 18-gauge needle at the 40% iodixanol layer. After washed with PBS, the viral fraction was stored at -80°C until the use for transfection.

Statistical analysis

Statistical significance tests were performed using SigmaPlot software (ver. 14.0) through a two-tailed Student's t-test or Welch's t-test. In case where normality fails, a Mann-Whitney Rank Sum test was employed. *P*-values <0.05 were considered significant. Data points in box and dot plots represent the full range of values, and boxes span the interquartile range (25th-75th percentile). Median and average values are indicated by horizontal lines. The error bars in all dot and bar plots show the standard deviation and were plotted with SigmaPlot (v. 14.0). We did not predetermine sample sizes based on statistical methods. For all experimental results, the n is reported in the accompanying figure legend.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Data that support the findings of this study are available within the Article and its Supplementary Information. Critical deep sequencing data were deposited at the NCBI Sequence Read Archive (SRA, http://www.ncbi.nlm.nih.gov/sra) under accession number PRJNA_. All other data that support the findings of the present study and plasmid vectors are available from the corresponding author upon request. Source data are provided with this paper.

Code availability

No computer codes were used in this study.

Figures

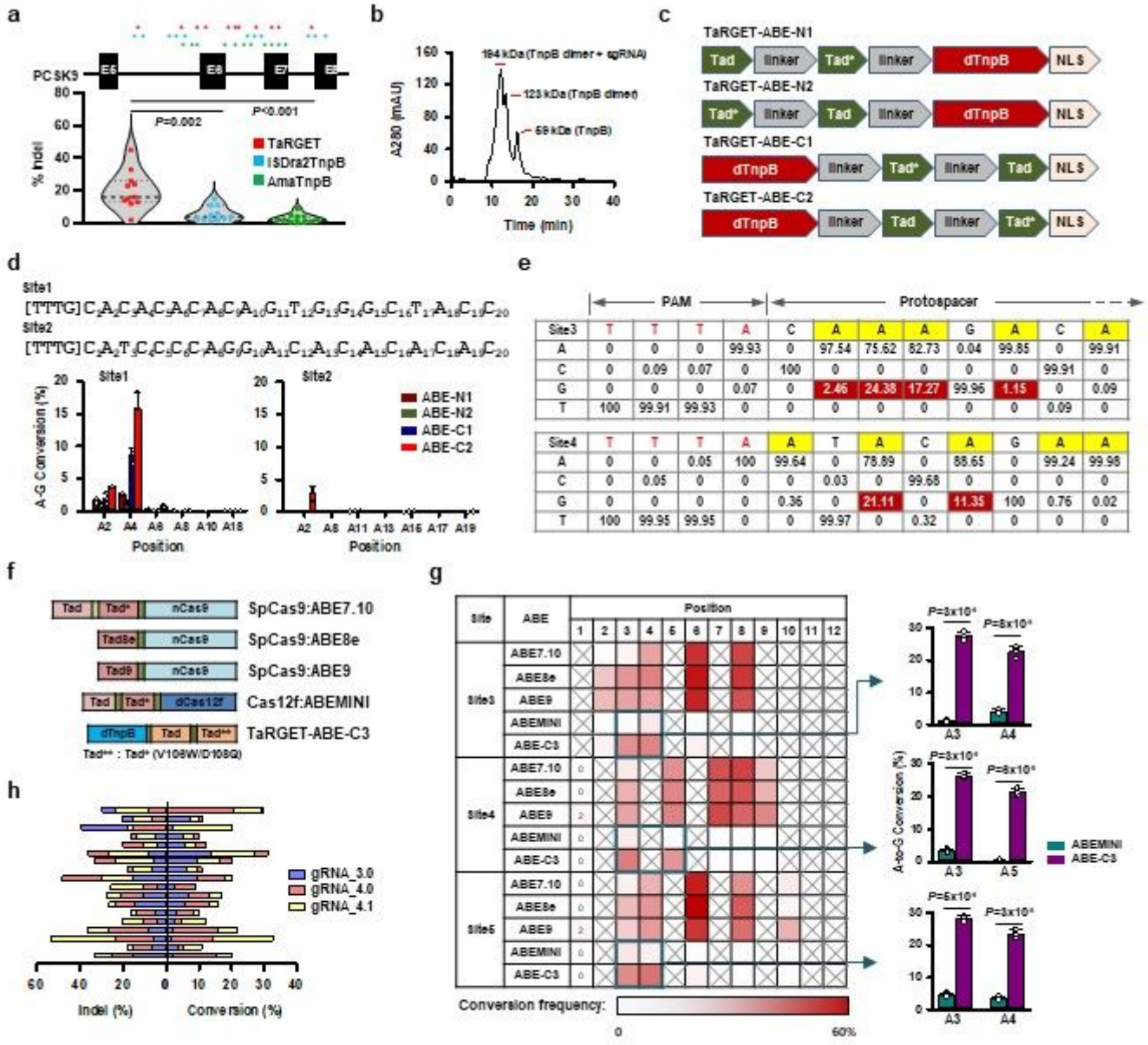

Figure 1

Figure 1

Feasibility test and optimization of the TaRGET-ABE system. a. Violin plots suggesting that the TaRGET system offers a favorable scaffold for adenine base editors due to the high indel efficiency, compared to other RNA-guided TnpB systems. n=11. *P*-values were derived by a two-tailed Welch's t-test (TaRGET versus ISDra2TnpB) and by the Mann-Whitney Rank Sum test (TaRGET versus AmaTnpB). **b**. Size-exclusion chromatography profiles of TnpB proteins in the presence of sgRNA. TnpB proteins were incubated with gRNA and the RNP complex was resolved on a Superdex 200 column. **c**. The structures and designations of TaRGET-ABE modules according to the orientation and order of the wild-type and mutant Tad. **d**. Investigation of the A-to-G conversion efficiency for two sites in HEK293T cells, each of which carries multiple A sequences at the PAM-proximal or –distal regions. n=3. **e**. The sequence profiles

at the PAM-proximal regions for two sites. Sequence profiles were obtained by deep-sequencing analysis of the sites following the transfection of the TaRGET-ABE-C2 vector. **f**. Architecture of SpCas9-, Cas12f1-, and TnpB-based adenine base editors. Tad** denotes the engineered form with V106W and D108Q mutations in Tad* protein. **g**. Heat map of the A-to-G conversion efficiency at three different sites for SpCas9-, Cas12f1-, and TnpB-based adenine base editors. X denotes the non-A sequence. In the panel on the right, the conversion efficiencies were compared between Cas12f-based ABEMINI and TaRGET-ABE-C3. n=3. *P*-values were derived by a two-tailed student's t-test. **h**. Dependence of indel and the A-to-G conversion efficiency on the types of engineered sgRNAs. n=18. All error bars represent the standard deviation (SD).

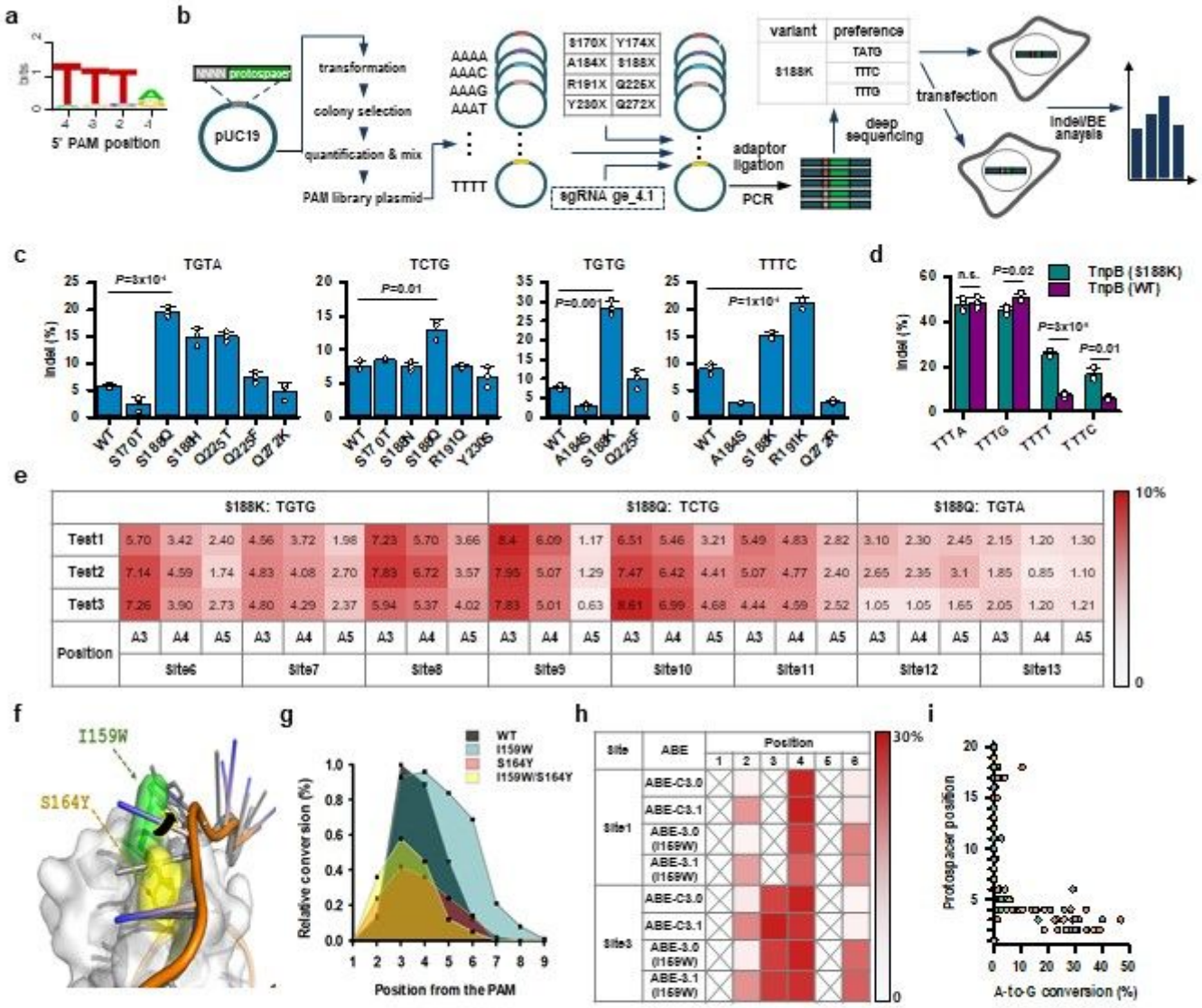

Figure 2

Figure 2

Expanding targetable sites by TnpB engineering and Tad optimization a. Sequence logo analysis of the PAM preference of the wild-type TnpB. **b**. Flow chart describing the development of PAM variants of TnpB and the application of the PAM variants to expanded adenine base editing. A validated target sequence in an NLRC4 *locus* (5'-TTTAGAGGGAGACACAAGTTGATA-3') was used throughout the experiments including *in vitro* cleavage and validation in PAM-altered cells. **c**. Investigation of the indel-forming activities of PAM variants for other PAM sequences than the canonical TTTR sequence. *P*-values were derived by a Student's test. n=3. **d**. Preference of the S188K TnpB variant for expanded PAM sequences (TTTN). n=3. **e**. Heat map showing the A-to-G conversion efficiencies of the PAM variants, S188K and S188Q TnpB PAM variants at non-TTTR PAM sites. Sites with As at positions 3, 4, and 5 were selected to trace the conversion efficiencies. Test 1-3 indicates triplicate experiments. **f**. Modeling of amino acid substitution for shifting or altering an editing window in the PAM-proximal region. An arrow indicates the effects of substituted bulky amino acids on the base protruding for easy access of deaminases. Ser164 and Ile159 were selected as the substitution candidates. **g**. Base editing window expanded by the substitution of Ile159 with tryptophan. Relative values were obtained by deriving the conversion efficiencies at each position for three different sites. n=3. **h**. Alteration of a base-editing window through Tad optimization. The deaminase composition of Tad-Tad8e (V106W, D108Q) extends the editing window to position 2. Values are the means of triplicate experiments. n=3. **i**. Multiple-site validation of TaRGET-ABE-C3.1 (V106W, D108Q) for A-to-G conversion activity. Values are the means from triplicate experiments. In total, 25 sites were investigated.

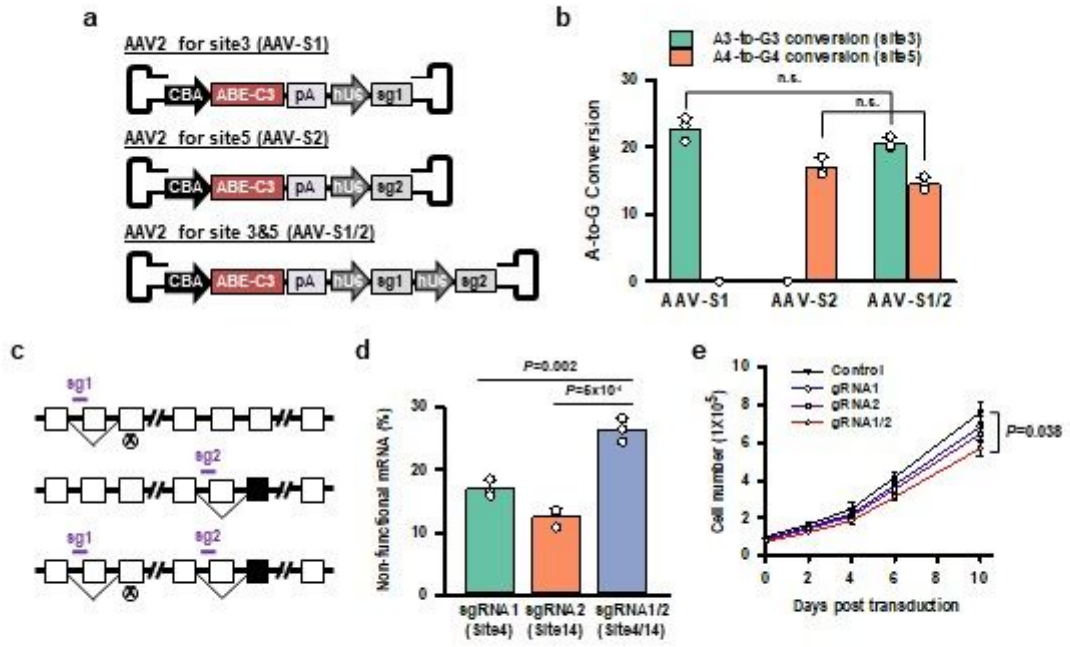

Figure 3

Figure 3

Validation of the A-to-G conversion activity of the TaRGET-ABE-C3 system through AAV delivery a. AAVs vector constructs harboring a TaRGET-ABE-C3 cassette and single or paired sgRNAs. **b**. Multiplexed adenine base editing by harboring paired sgRNAs in a single AAV particle. HEK293 cells ($1x10^5$ cells) were transduced at a multiplicity of infection (MOI) of 100,000 for an initial five days. Cells were splited five days after the onset of transduction and AAV2 particles (10^5 particles) were additionally added to the cells. P-values were derived by a Student's t-test. n=3. **c**. Schematic illustration showing the increased generation of non-functional mRNA using paired gRNAs. ⊗ denotes the occurrence of a non-sense mutation leading to an early termination of expression. ■ indicates a change of a functional exon into a non-functional one due to the deprivation of a former exon. **d**. Increased levels of non-functional ErbB4

mRNA by paired gRNAs in H661 cells. P-values were derived by a Student t-test. n=3. **e**. Growth inhibition of the TaRGET-ABE-C3 system harboring paired gRNA. Cells were transduced with AAV2 particles at an MOI of 10^5 for five days. Then, cells were sub-cultured at an initial cell number of $1x10^5$ and 10^5 AAV particles were again added to the cultured media. The cell number was counted from the day of sub-culturing for ten days. n=3. *P*-values were derived by a Fisher's least significant difference post hoc test. All error bars represent the SD.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- KimSupplementaryTable1.xlsx
- KimSupplementaryTable2.xlsx
- KimSupplementaryTable3.xlsx
- SupplementaryFigures.pptx
- ReportingSummary.pdf.pdf